

ARMY RESEARCH LABORATORY



# Document Classification by Fuzzy Attribute Evaluation

Aivars Celmiņš

ARL-TR-380

December 2000

Approved for public release; distribution unlimited.

20010122 138

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

# Army Research Laboratory

Aberdeen Proving Ground, MD 21005-5066

---

ARL-TR-380

December 2000

## Document Classification by Fuzzy Attribute Evaluation

Aivars Celmiņš

Computational and Information Sciences Directorate

---

Approved for public release; distribution unlimited.

---

---

## Abstract

---

A technique for classifying objects is proposed that combines classification estimates based on the properties of object attributes. The technique was developed as an aid to classifying captured foreign documents on the battlefield. In this approach, the input information consists of linguistic assessments of the document's classification; these assessments are based on document attributes such as document age, format, and place of discovery. The assessments are modeled as fuzzy sets and combined with the help of a decision function into an output fuzzy set that represents the overall assessment of the document. For a final linguistic classification, the output of the decision function is compared with target classes. The procedure achieves good performance if the decision function is trained on representative sets of classified objects.

Although the classification procedure was developed for classification of captured documents, it might be also applied to target recognition from approximate sensor inputs, triage procedures and diagnostics in medical praxis, risk assessments, and similar problems where classification requires the combination of uncertain judgments.

---

## Contents

---

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>                                     | <b>1</b>  |
| <b>2</b> | <b>Attribute Evaluation</b>                             | <b>3</b>  |
| 2.1      | Document Attributes . . . . .                           | 3         |
| 2.2      | Linguistically Scaled Significance Categories . . . . . | 4         |
| <b>3</b> | <b>Significance Accumulation</b>                        | <b>5</b>  |
| 3.1      | Decision Function . . . . .                             | 5         |
| 3.2      | Target Categories . . . . .                             | 6         |
| 3.3      | Proximity Indicators . . . . .                          | 7         |
| 3.4      | Decision Function Training . . . . .                    | 8         |
| 3.5      | Generation of Training Sets . . . . .                   | 9         |
| 3.6      | Linguistic Interpretation . . . . .                     | 11        |
| <b>4</b> | <b>Example</b>  | <b>13</b> |
| 4.1      | Training and Test Sets . . . . .                        | 13        |
| 4.2      | Training Results . . . . .                              | 13        |
| <b>5</b> | <b>Conclusions</b>                                      | <b>16</b> |
|          | <b>References</b>                                       | <b>17</b> |
|          | <b>Distribution</b>                                     | <b>19</b> |
|          | <b>Report Documentation Page</b>                        | <b>21</b> |

## Figures

|   |  |    |
|---|--|----|
| 1 | Membership functions of linguistic significance categories . . . | 4  |
| 2 | Membership function changes when parameters are modified .       | 6  |
| 3 | Effects of weight parameters on decision function . . . . .      | 7  |
| 4 | Membership functions of captured document categories . . . .     | 7  |
| 5 | Data set with sample correlation 0.9 . . . . .                   | 10 |

|   |  |    |
|---|--|----|
| 6 | Example of a decision process . . . . .    | 12 |
| 7 | Training history . . . . .                 | 14 |
| 8 | Classification of a training set . . . . . | 15 |
| 9 | Classification of a test set . . . . .     | 15 |

## Tables

|   |                               |   |
|---|-------------------------------|---|
| 1 | Document attributes . . . . . | 3 |
|---|-------------------------------|---|

---

## 1. Introduction

---

Captured foreign documents can be an important source of vital information on the battlefield. The handling of such documents is regulated by Army Field Manual 34-52 [1] (pp 4-2 to 4-4), which stipulates that captured documents should be assigned to one of four categories (named A, B, C, and D), depending on the contents of the documents, and that the documents should be dealt with according to their categories. Most important are documents of category A, which require immediate action; documents of categories B and C are less important, and documents of category D can be discarded.

A problem with this classification can arise when the documents are not in English, because then the contents of captured documents might not be obvious. Some help is provided by the FALCON system [2], which scans the documents and provides a quick translation and a simple computer analysis of the translated text (mainly by keyword searching). The soldier can use this analysis to assess the importance of the documents and eventually to categorize the documents.

Two improvements to the system are being considered. First, it is proposed to analyze the original text instead of its English translation [3]. Second, the input information for classification decisions is being extended beyond text analysis to other document attributes, such as document date, circumstances of capture, document type, etc. Under ideal circumstances, when the documents are in English, such additional attributes need not be considered, because an understanding of the text outweighs all other information. However, additional document attributes can be important when the contents of the documents are available only through a cursory computer analysis. Therefore, for an automatic document classification, particularly when the source language is not English, the input from nontextual attributes should be fused with outputs from computer text analysis. This report addresses the incorporation of nontextual document assessments into the classification procedure.

In the approach presented here, the importance of each document is expressed by a numerical significance value that is based on the combined significance indicators obtained from the attributes of the document. The attributes can be either outputs from text analysis or the other document properties mentioned above, and the significance indicator values are combined by a decision function that computes an accumulated significance value. The significance indicators can be vague linguistic expressions (such as "medium importance," "low importance," etc). These expressions, as well as outputs from text analysis, are modeled by fuzzy sets on a significance scale from zero to unity. The decision function accumulates these

fuzzy sets by calculating an output fuzzy set that represents the significance of the document. Finally, the accumulated significance is compared with standard significance categories (for instance, the Army's A, B, C, and D categories) and the result interpreted and formulated in linguistic form, such as "the document is approximately category A," "the document is likely secret," etc.

Section 2 describes common document attributes and corresponding significance indicators. Section 3 treats the computation and training of the decision function, and section 4 provides examples of document classification by attributes. A summary and conclusions are given in section 5.



## 2. Attribute Evaluation

### 2.1 Document Attributes

Table 1 gives a tentative list of attributes that might contribute to document classification and can be assessed without an understanding of document content. (Eventually, actual documents will be used to establish a more comprehensive list.) The assumption is that to obtain the significance level of a document, one would inspect the document and estimate values of significance indicators for all available attributes. Each "significance indicator" is a linguistic assessment of document significance derived from properties of an attribute; the value of the indicator is assessed for individual attributes *independently of the properties of other attributes*. For each attribute, the table lists some properties that might be used in estimating the value of the significance indicator. For example, the value of the indicator for the attribute "age of document" might be estimated as "medium high" if the document is current, and "low" if the document is several months old. Note that any attribute can have an entry equivalent to "unknown," which is treated as "not significant." Such an entry does not affect the accumulated document significance value.

Table 1. Document attributes.

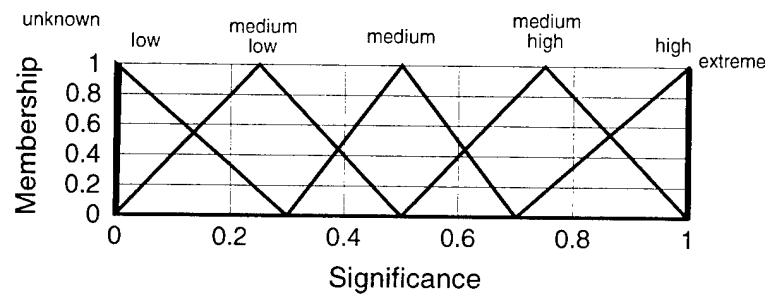
| Attribute                 | Indicator              | Attribute                         | Indicator                                  |
|---------------------------|------------------------|-----------------------------------|--|
| <b>Age of document</b>    | Current (present date) | <b>Style of document</b>          | Typed                                      |
|                           | Recent (few days old)  |                                   | Handwritten                                |
|                           | Weeks                  |                                   | Printed                                    |
|                           | Months                 |                                   | Fax  |
|                           | Unknown                |                                   | Carbon, photographic, or similar copy      |
| <b>Format of document</b> | Military order         | <b>Circumstances of discovery</b> | Troop quarters                             |
|                           | Report                 |                                   | Abandoned house                            |
|                           | Letter                 |                                   | Office                                     |
|                           | Indistinct             |                                   | File cabinet                               |
| <b>Stationery</b>         | Military stationery    |                                   | Open field                                 |
|                           | Business stationery    |                                   | On a person                                |
|                           | Loose leaf             | <b>Text analysis*</b>             | Frequency of military keywords             |
|                           | Notebook               |                                   | Military acronyms                          |
|                           |                        |                                   | Standard military expressions              |
|                           |                        |                                   | Frequency of a particular keyword class    |
|                           |                        |                                   | Text arrangement in military standard form |

\*There might be several indicators for this group, depending on the sophistication of the text analysis program. The list is a short tentative selection of possible subgroups.

## 2.2 Linguistically Scaled Significance Categories

Values of the significance indicators are assessed in linguistic terms. For this I use a five-category scale of fuzzy numbers from Chen and Hwang [4] (p 468) with the categories "low," "medium low," "medium," "medium high," and "high." The membership functions of these numbers are shown in figure 1. For present purposes, the five-category scale is supplemented with two extreme categories, which correspond to "unknown significance" and "extremely significant." The former is represented by a crisp singleton at  $s = 0$  (zero significance), and the latter is a crisp singleton at  $s = 1$ . The result is thus a scale of seven categories, a range that knowledge engineers consider optimal for linguistic estimates [4].

Figure 1. Membership functions of linguistic significance categories.



---

### 3. Significance Accumulation

---

#### 3.1 Decision Function

Let  $\tilde{a}_i$  be the value of the significance indicator from attribute  $i$ . The indicators are crisp or fuzzy numbers between zero and unity. An overall significance level for the document is obtained by the accumulation of the  $\tilde{a}_i$  with the help of a decision function. Let  $n$  be the number of attributes of a given document. Then the accumulated significance  $\tilde{S}$  of that document is computed by

$$\tilde{S} = 1 - \left[ \prod_{i=1}^{i=n} (1 - \langle m_i \cdot \tilde{a}_i \rangle)^{w_i} \right]^{0.5[1 + \exp(0.1 \cdot (1-n))]} \quad (1)$$

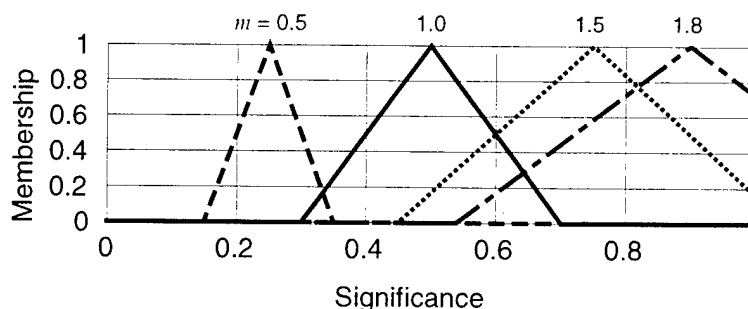
This decision function formula contains two sets of crisp parameters: attribute modifiers  $m_i$  and attribute weights  $w_i$ . These parameters make the decision function adaptable to particular applications. Their values are determined by a training procedure performed on representative sets of classified documents (see sect. 3.4).

The modifiers  $m_i$  are positive and enter the formula for  $\tilde{S}$  as crisp multipliers of the fuzzy indicators  $\tilde{a}_i$ . (For numerical reasons, the lower bound of the multipliers is set equal to 0.001.) The purpose of the modifiers is to change the fuzzy significance indicators by increasing or decreasing their values. Because the significance measure  $s$  is restricted to the interval  $[0,1]$ , a special truncating multiplication is used (instead of an ordinary multiplication) and indicated in equation (1) by  $\langle m_i \cdot \tilde{a}_i \rangle$ . That multiplication truncates the product to the unit interval and assigns to the abscissa  $s = 1$  a membership value that equals the maximum membership of those parts of the product that have an abscissa larger than one. For instance, if the multiplication by a large factor  $m_i$  shifts the support of  $\tilde{a}_i$  completely out of the unit interval, then the product  $\langle m_i \cdot \tilde{a}_i \rangle$  is a crisp singleton at the significance level  $s = 1$  (linguistically, "extremely significant").

The effects of the modifying parameters on significance membership functions are illustrated by figure 2. The original membership ("medium significance") is labeled with  $m = 1$ . If the modifying parameter is less than unity, the membership function is shifted to the left, and its spread is reduced. For modifying parameters larger than unity, the membership function is shifted to the right, its spread increases, and eventually (for large  $m$ ) it is shifted out of the unit interval. In that case, the product equals a singleton at  $s = 1$ .

The weights  $w_i$  are restricted to values larger than or equal to 0.1 and enter the formula as exponents of the contributions from the attributes  $i$ . A weight  $w_i$  has in essence the same effect as a  $w_i$ -times repetition of the  $i$ th

Figure 2. Membership function changes when parameters are modified.



estimate  $\hat{a}_i$ . Figure 3 shows the effect of weight parameters on a decision function, for simplicity assuming only one attribute and a modifying parameter  $m_1 = 1$ . If the weight parameter  $w_1 = 1$ , the membership function of  $\tilde{S}$  equals the input membership, labeled with  $w = 1$ . If the weight parameter is less than unity, the input membership function is shifted to the left, and its spread is reduced. Hence, its effect is similar to that of a small modifying parameter. For this reason, to avoid redundant parameter adjustments, I set the minimum permissible value of  $w_i$  to 0.1 (instead of zero). For  $w_i > 1$ , the membership of the decision function is shifted to the right and its spread reduced. It never shifts out of the unit interval. This is different from the effects of the modifying parameters, and it adds flexibility to the decision function.

The exponent of the square brackets in equation (1) has been determined experimentally. It reduces the contributions of individual significance indicators when the number  $n$  of attributes is large. The accumulated significance  $\tilde{S}$  of the document is a fuzzy set with support between zero and unity on the significance scale  $s$ .

### 3.2 Target Categories

The purpose of document classification is to assign each document to a category from a predefined set of target categories. In the approach presented here, the assignments are done by comparison of the accumulated significance  $\tilde{S}$  (a fuzzy set) with fuzzy sets that represent the target categories.

Target categories are in this case the four captured document categories defined in the relevant Army Field Manual [1]. These categories are modeled as fuzzy sets over the significance scale shown in figure 4. For consistency with the set of input categories, the set is supplemented with two extreme categories,  $D-$  and  $A+$ , which are represented as singletons at  $s = 0$  and  $s = 1$ , respectively. This is an arbitrary representation of the Army's categories, because the latter are defined in terms of the textual contents of the documents and not by their significance level. However, because the analysis presented here is based on fuzzy-set representation and the decision function is determined by a training procedure, a fuzzy-set representation of the target categories is adequate. It suffices that the membership functions of the four categories are in ascending order; the details of the functions are less important.

Figure 3. Effects of weight parameters on decision function.

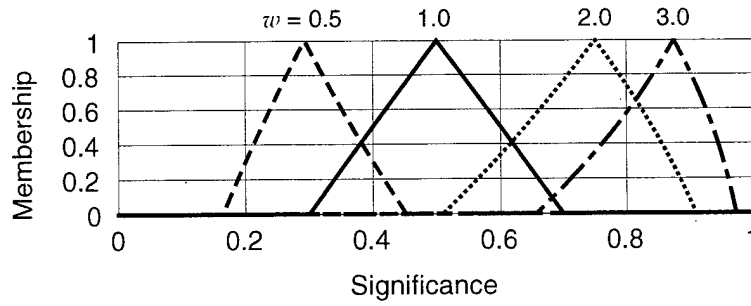
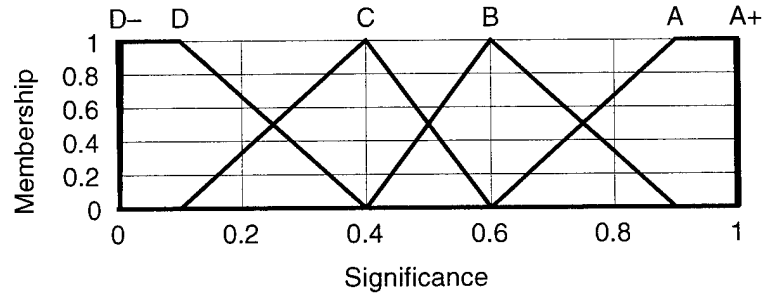


Figure 4. Membership functions of captured document categories.



### 3.3 Proximity Indicators

To train the decision function, one needs a measure of the goodness of document classification, that is, a measure of the deviation of a category implied by the decision function from the known category of a document. Because the accumulated significance and the target categories are fuzzy sets, the goodness measure must be a measure of disparities between fuzzy sets. For present purposes, the disparity between an accumulated significance  $\tilde{S}$  and a target category  $\tilde{C}$  is expressed by three proximity indicators, which are referred to as separation, discord, and exclusion.

The *separation* is the difference between the defuzzified values of  $\tilde{S}$  and  $\tilde{C}$ , where the defuzzified values are computed by the center of gravity method. Let  $\mu_S(s)$  and  $\mu_C(s)$  be the membership functions of  $\tilde{S}$  and  $\tilde{C}$ , respectively, and  $G_S$  and  $G_C$  be the corresponding defuzzified values. The defuzzified value  $G_S$  (the center of gravity of  $\mu_S(s)$ ) is computed by

$$G_S = \frac{\int_0^1 s \mu_S(s) ds}{\int_0^1 \mu_S(s) ds} \quad (2)$$

and  $G_C$  is computed correspondingly. The separation of  $\tilde{S}$  from a target category  $\tilde{C}$  is defined by

$$P_{SC} = G_S - G_C. \quad (3)$$

The separation  $P_{SC}$  can have any value between -1 and 1, and its sign indicates whether  $\tilde{S}$  is mainly to the left or mainly to the right of  $\tilde{C}$ .

The *discord* [5] between the two fuzzy sets  $\tilde{S}$  and  $\tilde{C}$  is defined by

$$D_{SC} = 1 - \max_s [\min(\mu_S(s), \mu_C(s))]. \quad (4)$$

The discord varies between zero and unity. It equals zero when the cores of the two fuzzy numbers intersect, and it equals unity when their supports do not intersect. The discord expresses the lack of intensity of coincidence between the two fuzzy sets.

The *exclusion* is a measure for the lack of overlap of the membership functions  $\tilde{C}$  and  $\tilde{S}$ . The exclusion  $E_{C(S)}$  of a target category  $\tilde{C}$  from  $\tilde{S}$  is defined by

$$E_{C(S)} = 1 - \frac{\int_0^1 \min [\mu_S(s), \mu_C(s)] ds}{\int_0^1 \mu_C(s) ds}. \quad (5)$$

Note that the exclusion is not symmetric. It equals zero when the membership function  $\mu_C(s)$  of the target category  $\tilde{C}$  is contained entirely in the set  $\mu_S(s)$ . It equals unity when  $\tilde{S}$  and  $\tilde{C}$  do not intersect. Otherwise  $E_{C(S)}$  is positive and less than or equal to unity. If  $\mu_C(s)$  is a singleton at  $s = s_C$ , then the exclusion is defined by

$$E_{C(S)} = 1 - \mu_S(s_C). \quad (6)$$

(In this case, the exclusion and the discord are identical.)

For the training of the decision function, a *disparity* between two fuzzy sets  $\tilde{S}$  and  $\tilde{C}$  is defined by

$$d_{S(C)} = \left[ \frac{D_{SC}^2 + E_{C(S)}^2}{2} \right]^{1/2} \cdot P_{SC}. \quad (7)$$

The disparity  $d_{S(C)}$  is a crisp number between  $-1$  and  $+1$ .

### 3.4 Decision Function Training

The relative importance of attributes for document classification is modeled by the values of the decision function parameters  $m_i$  and  $w_i$ . To determine the proper values of these parameters, one would classify a set of documents with known significance levels and find such parameter values that the document set overall is correctly classified. Let the training set consist of  $k$  documents with significance categories  $\tilde{C}_j$ ,  $j = 1, \dots, k$ . Each document is also characterized by a set of significance indicators according to its attributes. The significance indicators are accumulated by the decision function (eq (1)), providing for each document a fuzzy document significance  $\tilde{S}_j$ . The goal of the training is to modify the decision function parameters such that the fuzzy sets  $\tilde{S}_j$  are close to the corresponding target sets  $\tilde{C}_j$ . As a measure of agreement, I use the sum  $U$  of squares of disparities between the accumulated significances and the target categories, where the disparities are computed with equation (7):

$$U(m_1, m_2, \dots, m_n, w_1, w_2, \dots, w_n) = \frac{1}{k} \sum_{j=1}^{j=k} \left( d_{S(C)}^{(j)} \right)^2. \quad (8)$$

In equation (8),  $d_{S(C)}^{(j)}$  is the disparity between the accumulated significance  $\tilde{S}_j$  and the target significance  $\tilde{C}_j$  for the document  $j$ . The training parameters are the attribute modifiers  $m_i$  and the weights  $w_i$ ; see equation (1). They are determined by a steepest descent algorithm on the objective function  $U$  in the parameter space. The partial derivatives of  $U$  that are needed for the steepest descent algorithm are numerically approximated by difference quotients.

### 3.5 Generation of Training Sets

In generating synthetic "documents" for use in training sets, one must model the attributes that characterize a given document set. The most useful attributes in this context are those that correlate highly with the classification of the document. For instance, assume that for a certain document type, two attributes (e.g., date and keywords) classify the document correctly most of the time. In this instance, the significance indicators from these two attributes should correlate with the document significance level. Therefore, training sets that represent that document type should be so constructed that sample correlation coefficients between the significance indicators of the two attributes and the corresponding document categories have values close to unity.

This section outlines the construction of training sets with prescribed correlations that can be used to test the training program. In real-life applications, a training set consists of documents that are considered typical for the relevant scenario.

Let the number of documents in the training set be  $k$ . Each "document" in the training set is represented by its target class  $C_j$ ,  $j = 1, \dots, k$  and a list of significance indicators  $\tilde{a}_i^{(j)}$ :

$$C_j, \tilde{a}_1^{(j)}, \tilde{a}_2^{(j)}, \tilde{a}_3^{(j)}, \dots, \tilde{a}_n^{(j)}.$$

For construction of the training sets, sample correlation coefficients  $\rho_i$ ,  $i = 1, \dots, n$  are prescribed between  $\tilde{a}_i^{(j)}$  and  $C_j$ .

The construction of training sets begins with a set of  $k$  crisp data points  $(x_j, y_j)$ ,  $j = 1, \dots, k$ . The sample correlation coefficient of the  $k$  data pairs  $(x_j, y_j)$  is defined by

$$\gamma_k(x, y) = \left( \sum_{j=1}^k (x_j - \bar{x})^2 \sum_{j=1}^k (y_j - \bar{y})^2 \right)^{-1/2} \sum_{j=1}^k (x_j - \bar{x})(y_j - \bar{y}), \quad (9)$$

where  $\bar{x}$  and  $\bar{y}$  are average values of the  $x_j$  and  $y_j$ , respectively. To obtain a set with a prescribed positive correlation  $\gamma_0$ , one can proceed as follows. First, two different values  $x_1$  and  $x_2$  are chosen at random, and corresponding  $y$ -values are set equal to the  $x$ -values:  $y_1 = x_1$  and  $y_2 = x_2$ . Next, the remaining  $x_j$  for  $j = 3, \dots, k$  are chosen at random, and the corresponding

$y_j$  are calculated such that the correlation

$$\gamma_j(x, y) = \left[ 1 - (1 - |\gamma_o|) \cdot (j/k) \right] \cdot \text{sgn}(\gamma_o). \quad (10)$$

The calculation is done by numerical search alternating above and below the line through the first two data points. If the prescribed  $\gamma_o$  is negative, then one can use the same algorithm by changing the initial  $y$ -values to  $y_1 = x_2$  and  $y_2 = x_1$ .

For this algorithm to be used for document generation, some adjustments are necessary. First, the values of  $x$  and  $y$  are restricted to the unit interval, and corresponding restrictions for the random choice and search algorithm apply. Second, for the initial data, one can use without loss of generality  $x_1 = 0.2$  and  $x_2 = 0.8$  instead of random values. Then only the subsequent  $x_j$  are chosen randomly from a uniform distribution over the unit interval. Finally, for generating randomized  $y_j$ , the algorithm for the computation of  $y_j$  was augmented as follows. Let  $\hat{y}_j$  be a value of  $y$  that produces the desired correlation coefficient  $\gamma_j$ , and let

$$Y_j = y_1 + (y_2 - y_1) \frac{x_j - x_1}{x_2 - x_1} \quad (11)$$

be the ordinate of the intersection point of the line  $x = x_j$  with the line through the first two data points. The value  $y_j$  is randomly chosen from an interval bounded by

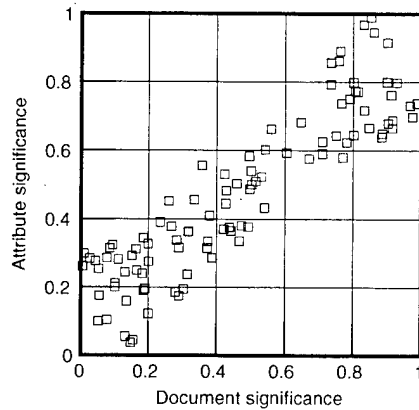
$$y_{\text{base}} = Y_j + (\hat{y}_j - Y_j) \cdot (j/k) \quad (12)$$

and

$$y_{\text{end}} = \min \left\{ 1, \max \left[ 0, y_{\text{base}} + 2(\hat{y}_j - y_{\text{base}}) \right] \right\}. \quad (13)$$

The interval shrinks as  $j$  increases, and it has zero length for the final data point with  $j = k$ , which ensures a correct final correlation coefficient for the set. Figure 5 shows a set with 100 data points, one attribute, and a sample correlation coefficient of 0.9, which was generated by the described algorithm.

Figure 5. Data set with sample correlation 0.9.





A conversion of the crisp data into fuzzy categories is needed to generate fuzzy significance estimates. For conversion (fuzzification) of the data, the  $y$ -axis was subdivided in segments corresponding to the linguistic categories (fig. 1) and the  $x$ -axis in segments corresponding to the Army's categories (fig. 4); fuzzy values  $\tilde{a}_i^{(j)}$  and  $C_j$  were assigned according to the compartment in which each point  $(x_j, y_j)$  was located.

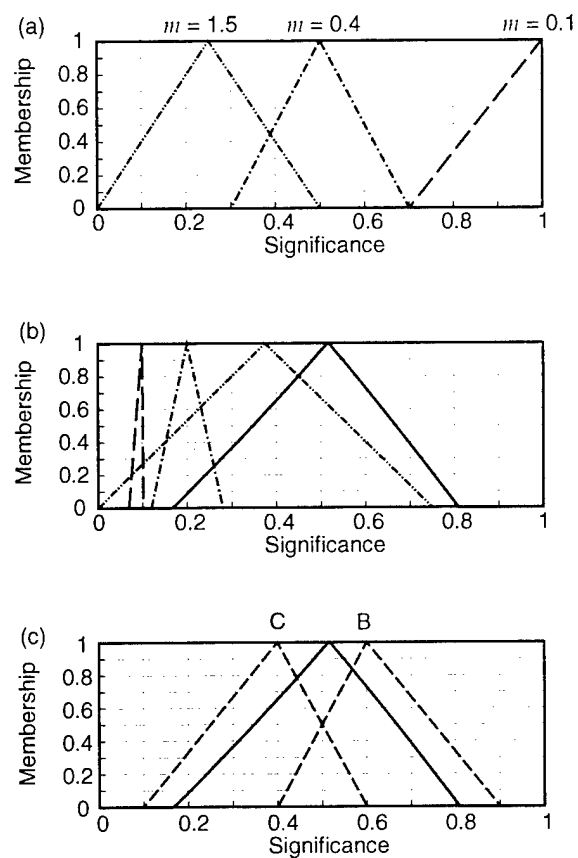
### 3.6 Linguistic Interpretation

The output of the decision function is a fuzzy set  $\tilde{S}$  on the significance scale. One obtains a linguistic interpretation of the result by comparing  $\tilde{S}$  with fuzzy sets that represent target categories. The comparison is done in terms of the disparity  $d_{S(C)}$  between fuzzy sets, and the document is assigned to the category that is closest to the accumulated significance in terms of the disparity. If  $\tilde{S}$  is between two target categories, then a corresponding hedge is added to the linguistic interpretation (for example, "with low confidence").

As an illustration of the classification procedure, consider a simple case with only three attributes of equal weights ( $w_i \equiv 1$ ) and modifying parameter values  $m_1 = 1.5$ ,  $m_2 = 0.4$ , and  $m_3 = 0.1$ . Let the corresponding significance indicators from the three attributes be "low medium," "medium," and "high." Figure 6(a) shows the input fuzzy sets. After multiplication with the parameters  $m_i$ , one obtains a modified set of three indicators, shown in figure 6(b). Accumulation of these three indicators by equation (1) produces a fuzzy set  $\tilde{S}$ , shown in figure 6(b) by a heavy solid line. Next,  $\tilde{S}$  is compared with target categories (see fig. 4); the comparison is shown in figure 6(c). Obviously,  $\tilde{S}$  is between categories B and C, and somewhat closer to B. The linguistically formulated result in this example is as follows:

*Document category is approximately B: Likely SECRET information with low confidence.*

Figure 6. Example of a decision process:  
 (a) input sets (line styles represent different sets);  
 (b) sets after multiplication (solid line is fuzzy set  $\tilde{S}$ ); and  
 (c) comparison of  $\tilde{S}$  (solid line) with target categories (dashed lines).



---

## 4. Example

---

### 4.1 Training and Test Sets

Consider an example of training that was performed on a training set of 100 synthetic "documents," generated as described in section 3.5. The number of attributes is  $n = 4$ , so that the input from each document consists of five items:

$$C_{\text{target}}, \tilde{a}_1, \tilde{a}_2, \tilde{a}_3, \tilde{a}_4.$$

The prescribed sample correlation coefficients  $\rho_i = \rho(C_{\text{target}}, \tilde{a}_i)$ ,  $i = 1, 2, 3, 4$ , were as follows:

$$\rho_1 = 0.9, \rho_2 = 0.9, \rho_3 = 0.1, \rho_4 = -0.9.$$

Accordingly, the first two attributes of the documents provide mostly correct information about document significance, while the information from attributes 3 and 4 is mostly false. We would therefore expect that, after the training, the first two attributes will be weighted more heavily than the last two.

A test set, also consisting of 100 "documents," was established in the same manner, with the only difference being that a different seed number was used in the random number generation routine.

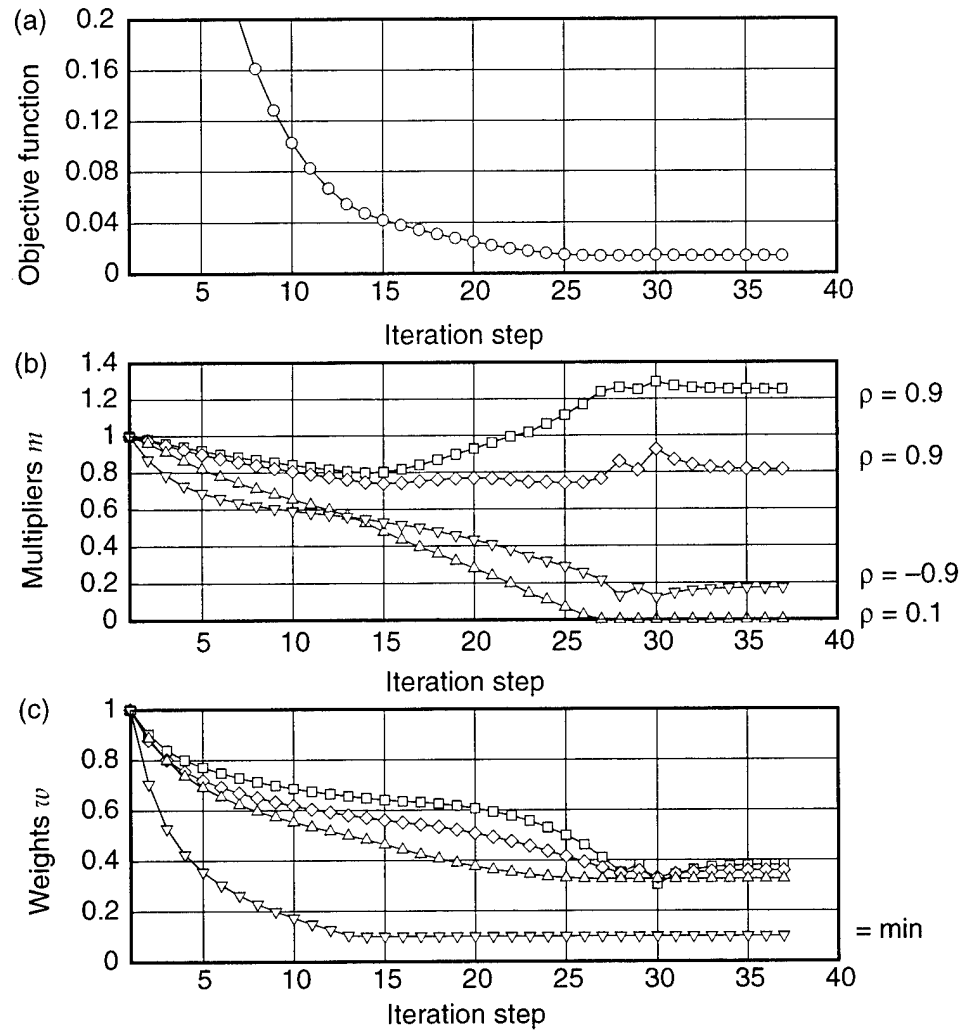
### 4.2 Training Results

The training of the decision function required 36 iterations. The training history is illustrated in figure 7. Figure 7(a) displays the value of the objective function  $U$  over the number of iteration steps. It shows that a minimum of  $U$  was found after about 29 iterations; the iteration was continued because of conservative iteration end conditions.

Figure 7(b) shows the development of the modifying parameters  $m_i$ . By the end of the iteration, the parameters of the "good" attributes,  $m_1$  and  $m_2$ , had settled to values of about unity. The "bad" attributes (3 and 4) both converged to very small values, making the contributions from these attributes negligible. The value  $m_3 = 0.001$  is the lower bound for the modifying parameters.

Figure 7(c) shows the development of the weight parameters  $w_i$ . The weights of the "good" parameters converge to values close to 0.4. The weight of  $w_3$  of attribute 3 has a similar value, but that attribute is already eliminated by its very small modifying parameter. The weight of attribute 4 converges to the smallest permissible value, 0.1. Because the parameters  $m_i$  and  $w_i$  have, for small values, similar effects on the decision function, an attribute can be eliminated by a small value of either parameter, and there

Figure 7. Training history: (a) value of  $U$  over iteration; (b) development of parameters  $m_i$ ; and (c) development of parameters  $w_i$ .



is no improvement from reducing the corresponding parameter from the other parameter set.

The final values of the parameters were as follows:

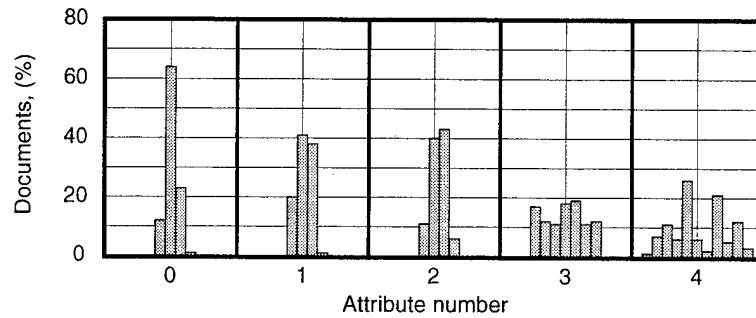
$$\begin{array}{llll} m_1 = 1.251 & m_2 = 0.813 & m_3 = 0.001 & m_4 = 0.173 \\ w_1 = 0.382 & w_2 = 0.360 & w_3 = 0.326 & w_4 = 0.100 \end{array}$$

In principle, attribute 4 with its large negative correlation could be used for classification if I inverted its significance indicator. However, the present setup of the decision function training does not allow such a usage: the model is based on the assumption that significance information from attributes is correct. If this is not the case for documents in the training set, then the corresponding attribute is suppressed.\*

The capability of the trained decision function to classify documents of the training set is illustrated in figure 8. This figure shows the distributions of

\*A modification of the decision function that allows it to handle negative correlations is presented elsewhere [6].

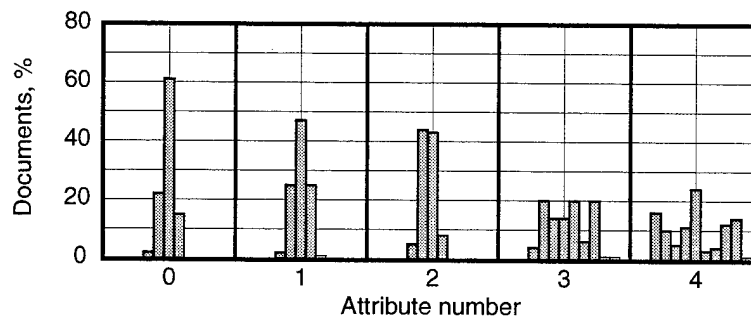
Figure 8. Classification of a training set.



classification errors by attribute significance indicators and, with the label "0," the classification errors by the decision function. For this display, a classification error is expressed in terms of the difference between the correct document class and the implied class by attributes and decision function, respectively. Because there are six target categories, a classification error is an integer with values between  $-5$  and  $+5$ . Accordingly, the error distributions are shown in 11 bins for each prediction. The figure shows that attributes 1 and 2 classify about 40 percent of the documents correctly and the remaining documents mostly with an error of  $+1$  or  $-1$  category. The classification errors of attribute 3 (with correlation 0.1) are randomly distributed, and the classification errors of attribute 4 (with correlation  $-0.9$ ) have a bimodal distribution. The distribution labeled "0" shows the distribution of classification errors by the trained decision function. It correctly classifies over 60 percent of the documents, which indicates that combining attribute significance indicators does indeed improve classification quality.

To test the success of the training, I prepared a test set of 100 documents with the same characteristics as the training set (i.e., with the same correlations between attribute significance indicators and document classes). The classification results for this set are shown in figure 9 in the same form as figure 8. The distributions of classification errors have a pattern similar to that in figure 8 and confirm the success of the training.

Figure 9. Classification of a test set.



---

## 5. Conclusions

---

A classification procedure has been devised in which attribute-provided classifications are combined with the help of a fuzzy decision function. The decision function contains parameters that are determined by the function being trained on representative training sets. The trained function classifies with fewer errors than classifications based on any single attribute. Because of the simplicity of the decision function and the transparency of the roles of the parameters, the training also provides indications about the importance of each attribute. Unimportant attributes are indicated by small parameter values and can be disregarded in subsequent classifications, thereby simplifying the classification process.

To improve the performance of the described classification method, an investigation of the effects of the decision function's structure would be helpful. In particular, the function parameters  $m_i$  and  $w_i$  might be supplemented or replaced by different parameters and the effects of such changes on classification performance studied.

The described classification procedure has potential applications well beyond the immediate use described in this report. Possible applications include document classification, target recognition, triage procedures, diagnostics, risk assessment, and other classification problems where the result depends on attribute properties.

---

## References

---

1. *Intelligence Interrogation*, Field Manual 34-52, Headquarters of the Department of the Army, Washington, DC (28 September 1992).
2. F. Fisher and C. R. Voss, "FALCON, An MT System Support Tool for Non-Linguists," *Proc. Advanced Information Analysis Conference*, McLean, VA (1997).
3. V. V. Cross and C. R. Voss, "Fuzzy Ontologies for Multilingual Document Exploitation," *Proc. 18th International Conference of the North American Fuzzy Information Processing Society (NAFIPS-99)*, pp 392–397, IEEE Publications, Piscataway, NJ (1999).
4. S.-J. Chen and C.-L. Hwang, *Fuzzy Multiple Attribute Decision Making*, Springer-Verlag, Berlin (1992).
5. A. Celmiņš, "Least Squares Model Fitting to Fuzzy Vector Data," *Fuzzy Sets Syst.* **22**, 245–269 (1987).
6. A. Celmiņš, "Classification with a Fuzzy Decision Function," *Proc. 8th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU)*, pp 267–270, Universidad Politécnica de Madrid, Spain (2000).

## Distribution

Admnstr  
Defns Techl Info Ctr  
Attn DTIC-OCF  
8725 John J Kingman Rd Ste 0944  
FT Belvoir VA 22060-6218

Ofc of the Secy of Defns  
Attn ODDRE (R&AT)  
The Pentagon  
Washington DC 20301-3080

Ofc of the Secy of Defns  
Attn OUSD(A&T)/ODDR&E(R) R J Trew  
3080 Defense Pentagon  
Washington DC 20301-7100

AMCOM MRDEC  
Attn AMSMI-RD W C McCorkle  
Redstone Arsenal AL 35898-5240

Dir for MANPRINT  
Ofc of the Deputy Chief of Staff for Prsnl  
Attn J Hiller  
The Pentagon Rm 2C733  
Washington DC 20301-0300

SMC/CZA  
2435 Vela Way Ste 1613  
El Segundo CA 90245-5500

US Army ARDEC  
Attn AMSTA-AR-TD  
Bldg 1  
Picatinny Arsenal NJ 07806-5000

US Army Info Sys Engrg Cmnd  
Attn AMSEL-IE-TD F Jenia  
FT Huachuca AZ 85613-5300

US Army Materiel Cmnd  
Attn AMCRD MG Caldwell  
5001 Eisenhower Ave  
Alexandria VA 22333-0001

US Army Natick RDEC Acting Techl Dir  
Attn SBCN-T P Brandler  
Natick MA 01760-5002

US Army Simulation, Train, & Instrmntn  
Cmnd

Attn AMSTI-CG M Macedonia  
Attn J Stahl  
12350 Research Parkway  
Orlando FL 32826-3726

US Army Tank-Automtv Cmnd Rsrch, Dev, &  
Engrg Ctr  
Attn AMSTA-TR J Chapin  
Warren MI 48397-5000

US Army Train & Doctrine Cmnd  
Battle Lab Integration & Techl Dirctr  
Attn ATCD-B  
FT Monroe VA 23651-5850

US Military Academy  
Mathematical Sci Ctr of Excellence  
Attn MADN-MATH MAJ M Huber  
Thayer Hall  
West Point NY 10996-1786

Nav Surface Warfare Ctr  
Attn Code B07 J Pennella  
17320 Dahlgren Rd Bldg 1470 Rm 1101  
Dahlgren VA 22448-5100

DARPA  
Attn S Welby  
3701 N Fairfax Dr  
Arlington VA 22203-1714

Univ of Texas at Austin  
Inst for Advncd Tchnlgy  
PO Box 202797  
Austin TX 78720-2797

Hicks & Associates Inc  
Attn G Singley III  
1710 Goodrich Dr Ste 1300  
McLean VA 22102

US Army Soldier & Biol Chem Cmnd  
Dir of Rsrch & Techlgy Dirctr  
Attn SMCCR-RS I G Resnick  
Aberdeen Proving Ground MD 21010-5423



## Distribution (cont'd)

US Army Rsrch Lab  
Attn AMSRL-CI-C R Hammell  
Attn AMSRL-CI-H C Zoltani  
Attn AMSRL-CI-LP (305)  
Attn AMSRL-IS-CI A Celmins (5 copies)  
Attn AMSRL-IS-CI B Broome  
Bldg 1116A  
Aberdeen Proving Ground MD 21005  
  
Director  
US Army Rsrch Ofc  
Attn AMSRL-RO-D JCI Chang  
PO Box 12211  
Research Triangle Park NC 27709

US Army Rsrch Lab  
Attn AMSRL-DD J M Miller  
Attn AMSRL-CI-AI-R Mail & Records Mgmt  
Attn AMSRL-CI-AP Techl Pub (3 copies)  
Attn AMSRL-CI-C C Voss  
Attn AMSRL-CI-C J Gowens  
Attn AMSRL-CI-C M Holland  
Attn AMSRL-CI-LL Techl Lib (3 copies)  
Attn AMSRL-IS-TP A Brodeen  
Attn AMSRL-IS-TP B Cooper  
Adelphi MD 20783-1197

| REPORT DOCUMENTATION PAGE   |   |  | Form Approved<br>OMB No. 0704-0188                        |  |
|---|---|--|---|--|
| Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.  |   |  |   |  |
| 1. AGENCY USE ONLY (Leave blank)  |   | 2. REPORT DATE<br>December 2000                            |   | 3. REPORT TYPE AND DATES COVERED<br>Apr 99 to Jan 00 |
| 4. TITLE AND SUBTITLE<br>Document Classification by Fuzzy Attribute Evaluation  |   |  | 5. FUNDING NUMBERS<br>DA PR: AH44<br>PE: 611102H44        |  |
| 6. AUTHOR(S)<br>Aivars Celmiņš  |   |  |   |  |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>U.S. Army Research Laboratory<br>Attn: AMSRL-CI-CT email: celms@arl.army.mil<br>Aberdeen Proving Ground, MD 21005-5066  |   |  | 8. PERFORMING ORGANIZATION<br>REPORT NUMBER<br>ARL-TR-380 |  |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)<br>U.S. Army Research Laboratory<br>Aberdeen Proving Ground, MD 21005-5066  |   |  | 10. SPONSORING/MONITORING<br>AGENCY REPORT NUMBER         |  |
| 11. SUPPLEMENTARY NOTES<br>ARL PR: 0TE320<br>AMS code: 611102H411   |   |  |   |  |
| 12a. DISTRIBUTION/AVAILABILITY STATEMENT<br>Approved for public release;<br>distribution unlimited.   |   |  | 12b. DISTRIBUTION CODE                                    |  |
| 13. ABSTRACT (Maximum 200 words)<br>A technique for classifying objects is proposed that combines classification estimates based on the properties of object attributes. The technique was developed as an aid to classifying captured foreign documents on the battlefield. In this approach, the input information consists of linguistic assessments of the document's classification; these assessments are based on document attributes such as document age, format, and place of discovery. The assessments are modeled as fuzzy sets and combined with the help of a decision function into an output fuzzy set that represents the overall assessment of the document. For a final linguistic classification, the output of the decision function is compared with target classes. The procedure achieves good performance if the decision function is trained on representative sets of classified objects.<br><br>Although the classification procedure was developed for classification of captured documents, it might be also applied to target recognition from approximate sensor inputs, triage procedures and diagnostics in medical praxis, risk assessments, and similar problems where classification requires the combination of uncertain judgments. |   |  |   |  |
| 14. SUBJECT TERMS<br>Fuzzy decision function, significance accumulation, fuzzy data fusion  |   |  | 15. NUMBER OF PAGES<br>28                                 |  |
|   |   |  | 16. PRICE CODE  |  |
| 17. SECURITY CLASSIFICATION<br>OF REPORT<br>Unclassified  | 18. SECURITY CLASSIFICATION<br>OF THIS PAGE<br>Unclassified | 19. SECURITY CLASSIFICATION<br>OF ABSTRACT<br>Unclassified | 20. LIMITATION OF ABSTRACT<br>UL                          |  |